Construct a gene-to-metabolite network to screen the key genes of triterpene saponin biosynthetic pathway in *Panax notoginseng*

Running title: Select the ginsenoside key genes in *P. notoginseng*

Ning Wang, Long Wang, Lianwen Qi *, Xu Lu*

State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, PR China

*Corresponding authors, Tel/Fax: +86-25-83271379. E-mail addresses: fleude@126.com (L.W. Qi), luxu666@cpu.edu.cn

Abstract

Triterpene saponins are main active constituents of *Panax notoginseng*. Metabolites profiling of twelve triterpene saponins were analysed by HPLC-MS in leaf, petiole and root extracts of *P. notoginseng*. Most of the PPT type saponins, except ginsenoside Re, were mainly distributed in roots, while PPD type saponins were detected among various tissues. The total content of PPD type saponins decreased in the order of leaf, petiole and root. The expression patterns of four key genes (*PnFPS*, *PnSQS*, *PnDS* and *PnSE*) in the triterpene saponin biosynthetic pathway were measured by RT-qPCR. All the four investigated genes showed high expression levels in leaf. A gene-to-metabolite network was constructed through canonical correlation analysis. The results indicated that the expression levels of *PnFPS*, *PnSQS*, *PnDS* and *PnSE* had high correlation with PPD type saponins ginsenoside Rb_2, Rb_3 and Rc, while *PnSQS* was also highly correlated with Rb_1. Combining metabolic profiling, RT-qPCR and gene-to-metabolite network, we inferred that the leaf of *P. notoginseng* were

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/bab.1580.

the main biosynthesis site of PPD type saponins Rb_2, Rb_3 and Rc. The contribution to the biosynthesis of ginsenosides Rb_2, Rb_3 and Rc was in the order of *PnSE* > *PnDS* > *PnSQS* > *PnFPS*. *PnSE* and *PnDS* should be the preferred targets to regulate the production of PPD type saponins Rb_2, Rb_3 and Rc in *P. notoginseng* by plant metabolic engineering.

Key words: *Panax notoginseng*, HPLC-MS, gene expression, principal component analysis, canonical correlation analysis

Abbreviations: HPLC-MS, high performance liquid chromatography-mass spectrometry; FPS, Farnesyl diphosphate synthase; SQS, squalene synthase; DS, dammarenediol Ⅱ synthase; SE, squalene epoxidase; RT-qPCR, real-time quantitative polymerase chain reaction; PPT, 20(S)-protopanaxatriol; PPD, 20(S)-protopanaxadiol

1. Introduction

Panax notoginseng (*P. notoginseng*) is a highly valued medicinal herb that has been used in China for more than 600 years [1]. It belongs to the Ginseng genus, along with the two renowned species *Panax ginseng* and *Panax quinquefolius*, and its main active constituents are triterpene saponins [2, 3]. A wide range of pharmacological activities of *P. notoginseng* has been reported, such as hemostatic, anti-oxidant, hepatoprotective and anti-cancer functions [4-7]. The root of *P. notoginseng* has been widely used in the treatment of coronary heart disease and cerebral vascular disease [8]. This crude drug is also a major component of

common household medicines such as Yunnan Baiyao for the treatment of trauma and compound Danshen dripping pill for angina, and is widely consumed in China and other Asian countries in the form of tea, powder and common botanical ingredients in functional foods [5, 9, 10].

As a perennial herbaceous plant, the growth and development of *P. notoginseng* from seed to a mature plant takes several years, and its roots, the main medicinal part, are conventionally harvested until the third growth year. In addition, the available planting area of *P. notoginseng* is heavily limited for the reason that it is extremely particular about the growth environment [11]. As a result, the yield of *P. notoginseng* is far from enough to meet the market demands. Therefore, an increasing number of researchers have focused their studies on the regulation of triterpene saponin biosynthesis in *P. notoginseng*.

During the past few decades, a series of triterpenoid biosynthesis key genes have been cloned and characterized in *P. notoginseng* (Fig. 1). The precursor isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are synthesized by both the mevalonic acid (MVA) pathway in the cytosol and the methylerythritol phosphate (MEP) pathway in the plastid [12]. Farnesyl diphosphate (FPP) is derived from DMAPP together with two molecules of IPP via the catalysis of farnesyl diphosphate synthase (FPS) [13-15]. The head-to-head linkage of two FPP by the enzyme squalene synthase (SQS) of *P. notoginseng* leads to formation of squalene, which is further oxidized to 2, 3-oxidosqualene mediated by squalene epoxidase (SE) [15-17]. In the following step, dammarenediol Ⅱ synthase (DS) catalyzes the cyclization reaction of 2, 3-oxidosqualene to form dammarenediol, providing tetracyclic skeletons for dammarane-type saponins [15,18]. Subsequently, dammarenediol can be transformed to protopanaxadiol and protopanaxatiol by P450-dependent monooxygenases [CYP450s, including PnPPDS and P6H (PnPPTS)]. Even though *P.*

notoginseng contains varied ginsenosides from those of *P. ginseng*, it has almost identical sequences of PgPPDS and PgPPTS [19-22]. Protopanaxadiol and protopanaxatiol can be transformed to diverse ginsenosides through UDP-glycosyltransferases (UGTs) and glycosyltransferases (GTs) [23]. Then, a UDP-glucose:ginsenoside Rd glucosyltransferase (UDRdPG) was identified in *P. notoginseng* [24, 25].

However, less comprehensive analysis is available regarding the contribution of the key genes towards triterpene saponin biosynthesis in *P. notoginseng*. As the accumulation and transportation of triterpene saponins has been in existence for several years in mature plants, the expression level of triterpene saponin biosynthesis pathway genes and the distribution of triterpene saponins have an indirect relationship. One-year-old *P. notoginseng* seedlings are the preferred material to analyse the contribution of triterpene saponin biosynthesis key genes, for there is little transportation of triterpene saponins existed. In this study, the content of twelve triterpene saponins in different parts of one-year-old *P. notoginseng* were quantitatively analysed using HPLC-MS. The transcription level of *PnFPS*, *PnSQS*, *PnDS* and *PnSE* that involved in triterpene saponin biosynthesis were analysed by RT-qPCR. A gene-to-metabolite network was further built to screen the key genes of triterpene saponin biosynthesis pathway using canonical correlation analysis. This study can provide effective targets to improve triterpene saponin yield by plant metabolic engineering in *P. notoginseng*.

2. Materials and methods

2.1 Plant material

The seeds of *P. notoginseng* were come from Wenshan Prefecture, Yunnan Province, China. The seeds were sown in the greenhouse of China Pharmaceutical University. Leaf, petiole

and root of one-year-old *P. notoginseng* were collected respectively, immediately frozen in liquid nitrogen and stored at -80 ^{o}C.

2.2 Standards and chemicals

Methanol was purchased from Hanbon Tech Co. (Jiangsu Province, China). Acetonitrile of HPLC grade was obtained from TEDIA Company (Fairfield, Ohio, USA). Formic acid (99%, HPLC grade) was from ROE (Newark, New Castle, USA). Distilled deionized water was purified with a Milli-Q water purification system (Millipore, Bedford, MA, USA). The reference standards were purchased from the National Institute for the Control of Pharmaceutical and Biological Products (Beijing, China): notoginsenoside R_1, ginsenosides Rg_1, Re, Rf, Rb_1, Rg_2, Rg_3, Rc, Rh_1, Rb_2, Rb_3, Rd and jujuboside B (internal standard).

2.3 Sample extraction

The samples were homogenized in liquid nitrogen, and the powder was vacuum-lyophilized for 48 h. The powder samples were extracted according to the method developed by Lai et al. [26] with minor revision. The power samples (50 mg/sample) were extracted with methanol (1.5 mL) by vortexing for 1min followed by ultrasonication at room temperature for 30 min. The resultant mixture was allowed to stand at room temperature for overnight and then centrifuged at 13000 rpm for 10 min. All samples were filtered through 0.22 μm membrane filters before HPLC-MS analysis.

2.4 Metabolites analysis by HPLC-MS

The samples were quantitatively analysed using Agilent 1100 series LC/MSD trap system (Agilent Technologies, Santa Clara, California, USA) equipped with an electrospray

ionization (ESI) interface. The method of metabolites analysis was performed as described previously with some modifications by Li et al. [27].The chromatographic separation was conducted with an Agilent Zorbax SB-C18 column (4.6 mm × 250 mm, 5 μm) at 30 ^{o}C, with an injection volume of 10 μL. The mobile phase consisted of 0.1% formic acid water (A) and 0.1% formic acid acetonitrile (B). The mobile phases were programed as follows: 25% B at 0–5 min; 25–32% B at 5–10 min; 32% B at 10–35 min; 32–37% B at 35–40 min; 37% B at 40–44 min; 37–75% B at 44–45 min and 75% B at 45–55 min. The flow rate was set at 1 mL/min. The mass spectrometer was operated in the negative mode. Drying gas flow rate, 9 L/min; Capillary voltage, 3500 V; Fragmentor voltage, 120 V. Quantitative analysis was performed using selective ion monitoring (SIM) mode. The experiments were repeated in three biological replicates.

2.5 Data analysis

The ESI^{-} raw data were analysed by the Markerlynx applications manager version 4.1 (Waters, Manchester, UK) with the parameters set according to our laboratory custom standards. The data were further analysed by SIMCA-P software 11.5 (Umetrics, Umea, Sweden) for unsupervised principal component analysis (PCA). Before analysis, raw data were standardized in a digital text (.txt) file.

2.6 RNA isolation and quantitative real-time PCR (RT-qPCR)

Total RNAs were isolated using a RNA prep pure plant kit (BioTeke, Beijing, China) as described by manufacturer. 1 μg of high quality total RNA was reverse-transcribed to first-strand cDNA using random hexamer primers. RT-qPCR was performed on a LightCycler® 96 Real-Time PCR System (Roche) with SYBR Premix ExTaq^{TM} (Takara)

following the instruction. The thermal cycle conditions used were 2 min preincubation at 95°C, followed by 45 cycles of amplification (10 s at 95°C, 15 s at 57°C and 30 s at 72°C). The *Actin* gene was served as reference and relative gene expression was calculated with the comparative Ct method [28]. The experiments were repeated in three biological replicates. The primers used in RT-qPCR were listed in Table 2.

2.7 Construct a gene-to-metabolite network by canonical correlation analysis

The expression levels of triterpene saponin biosynthetic genes and the triterpene saponins content in different tissues of one-year-old *P. notoginseng* were correlatively analysed by canonical correlation analysis using Pearson's correlation coefficient [29, 30]. The gene-to-metabolite network was constructed to identify key genes that contributed greatly to triterpene saponins biosynthesis in *P. notoginseng*. The canonical correlation variables (U and V) showed correlation between gene expression and metabolites accumulations. The cutoff value of variable correlation coefficient was 0.5.

3. Results

3.1 Quantitative analysis in different tissues of one-year-old *P. notoginseng*

The leaf, petiole and root of one-year-old *P. notoginseng* were analysed by HPLC-MS. Twelve triterpene saponins were identified by comparing the retention time and mass spectra with standards. Fig. 2 and Supplement Fig. 1 show the chromatograms of different tissues in one-year-old *P. notoginseng* and the mixed standard solution. The concentration of each investigated saponin was calculated according to the calibration curves with the internal standard method.

Table 1 shows the content of the detected triterpene saponins in different tissues of *P. notoginseng*. Significant differences exist in the contents of 12 triterpene saponins among different tissues. The total saponin content decreased in the order of the leaf, root and petiole, with the leaf showed 3-fold and 8.5-fold higher than root and petiole, respectively. The most abundant triterpene saponin in leaf is ginsenoside Rb_3, followed by Rb_2 and Rc, their combined content accounts for >90% of the total triterpene saponins. However, the contents of the three saponins in root and petiole are much lower, accounting for less than 10% of the total saponins. The most abundant saponin in root is ginsenoside Rg_1, and it cannot be quantified in leaf or petiole. The content of ginsenoside Rb_1 varied slightly among different parts of one-year-old *P. notoginseng*.

3.2 Principal component analysis (PCA)

For analysis of the distribution and contents of saponins in different parts of one-year-old *P. notoginseng*, an unsupervised clustering method (PCA) was applied [31]. As shown in Fig. 3A, the scores displayed in the PCA plot could be readily divided into three relative clusters: leaf, petiole and root. The occurrence and content of triterpene saponins were significantly distinct in different parts of one-year-old *P. notoginseng*. Based on PCA, the first two principal components (PCs) explained 89.54% of the total variance. The first principal component (t1) explained 68.04% of the total variance, and the second principal component (t2) explained 21.50% (Fig. 3A). In PC1, the concentrations of R1, Rg_1, Rg_2, Rh_1 and Re

describe 63.53% of the total variance. The second PC (PC2) is described by R1, Rb_2, Rb_3, Rc and Rd, with 81.60% of the total variance (Fig. 3B).

3.3 Gene expression pattern analysis

RT-qPCR was used to analyse the expression patterns of *PnFPS*, *PnSQS*, *PnDS* and *PnSE* in one-year-old *P. notoginseng*. The gene expression level was calculated relative to that in leaf. Fig. 4 showed the expression levels of these four genes in root, petiole and leaf of one-year-old *P. notoginseng*. *PnDS* and *PnSE* were highly expressed in leaf, but scarcely expressed in petiole and root. *PnSQS* showed high expression in leaf and petiole. *PnFPS* had higher expression level in leaf and root than in petiole.

3.4 Construct a gene-to-metabolite network to screen the key genes of triterpene saponins biosynthesis pathway

The expression patterns of four triterpene saponin biosynthetic genes and the content of seven main triterpene saponins in different tissues of one-year-old *P. notoginseng* were used to construct a gene-to-metabolite network by canonical correlation analysis using Pearson's correlation coefficient (Fig. 5). U and V is a pair of canonical correlation variables in figure 5. The first pair of canonical correlation variables (U and V) showed significant correlation between the four genes' expression levels and metabolites' accumulation with a coefficient of 1. The cutoff value of variable correlation coefficient was 0.5. For example, variable correlation coefficient between *SE* expression levels and seven triterpene saponins contents were -0.415, -0.438, 0.215, 0.473, 0.980, 0.981 and 0.988, respectively. The values suggested that the expression level of *SE* was correlated with ginsenoside Rc, Rb_2 and Rb_3, but not or

scarcely correlated with the other four saponins. The results of canonical correlation analysis were summarized as follows: (1) *PnFPS*, *PnSQS*, *PnDS* and *PnSE* showed high correlation with PPD type saponins ginsenoside Rc, Rb_2 and Rb_3, while *SQS* was also highly correlated with PPD type saponin Rb_1. (2) The contribution to the biosynthesis of ginsenoside Rb_2, Rb_3 and Rc was in the order of *PnSE* > *PnDS* > *PnSQS* > *PnFPS*. (3) The investigated genes in this study showed no obvious correlation with PPT type saponins Rg_1, Re and Noto R1.

4. Discussion

Wan et al. reported that the underground parts of perennial *P. notoginseng* contained rich PPT and PPD type saponins, while aerial parts contained PPD type saponins only in perennial *P. notoginseng* by HPLC-ELSD [32]. In our study, the distribution of saponins in one-year-old *P. notoginseng* seedings showed high consistency with their study. The root contained both PPT and PPD type saponins, while the leaf and petiole mainly contained PPD type saponins with trace amount of PPT type ginsenoside Re by HPLC-MS analysis. Except ginsenoside Re, no PPT type saponins were detected in leaf and petiole of one-year-old *P. notoginseng*. Compared with HPLC-ELSD, HPLC-MS has higher sensitivity, which resulted in the detection of the trace amount PPT type ginsenoside Re in aerial parts. Notoginsenoside R1, the characteristic PPT type saponin constituent of *P. notoginseng* [33, 34], was only detected in the root of *P. notoginseng*, too. Based on the above experimental results, we speculate that PPT type saponins should be synthesized primarily in the underground part, and the key GTs of PPT type saponins should be highly expressed in the underground part of *P. notoginseng*. In one-year-old *P. ginseng* seedings, the content of PPD type saponin Rb_1 and Rd, and PPT type saponin Rg_1 and Re were determined, too [23]. The root of *P. ginseng* contained similar metabolites, including PPT type saponins Rg_1, Rd and PPD type saponins Rb_1. However, the leaf and petiole of *P. ginseng* also contained PPT type saponins Rg_1. The results showed that

although *P. ginseng* and *P. notoginseng* had a close genetic relationship, there were significant differences in the distribution of metabolites.

Niu et al. reported the expression levels of triterpene saponin biosynthetic pathway key genes *PnFPS*, *PnSQS*, *PnDS* and *PnSE* in different tissues of four-year-old *P. notoginseng* [15]. In our study, the expression patterns of *PnSQS*, *PnDS* and *PnSE* in different tissues of one-year-old *P. notoginseng* were similar to the above mentioned results, but the expression pattern of *PnFPS* was different. The investigated genes had high expression levels in the leaves, while PPD type saponins were mainly distributed in the leaves, too. The constructed gene-to-metabolite network further indicated that *PnFPS*, *PnSQS*, *PnDS* and *PnSE* had obvious correlation with PPD type saponins, but showed no significant correlation with PPT type saponins. All above results implied that PPD type saponin should be mainly biosynthesized in leaf of *P. notoginseng.*

Integration of transcripts and metabolites data to construct a gene-to-metabolite network is crucial for the regulation of secondary metabolism in plants [35]. The gene-to-metabolite network of *P. notoginseng* showed that the contribution to the biosynthesis of ginsenoside Rb_2, Rb_3 and Rc was in order of *PnSE* > *PnDS* > *PnSQS* > *PnFPS*. The contribution of *SE* and *DS* were far greater than that of *FPS* and *SQS* in the biosynthesis of PPD type saponins. Therefore, *SE* and *DS* should be the preferred targets to regulate the production of PPD type saponins Rb_2, Rb_3 and Rc in *P. notoginseng*. PPT type saponins were mainly distributed in root, while the investigated four genes were highly expressed in leaf. The gene-to-metabolite network showed that there was no significant correlation between the biosynthesis of PPT type saponins and the expression of four genes. So, these four investigated genes shouldn't be the suitable targets to regulate the production of PPT type saponins Rg_1, Re and Noto R1 in *P. notoginseng*.

5. Conclusion

In this study, the metabolic profile of twelve triterpene saponins was analysed in different tissues of one-year-old *P. notoginseng* by HPLC-MS. The root contains both PPT and PPD type saponins, while the leaf and petiole mainly contains PPD type saponins, with minor amount of ginsenoside Re. The expression patterns of *PnFPS*, *PnSQS*, *PnDS* and *PnSE* that involved in the triterpene saponin biosynthetic pathway were investigated by RT-qPCR. All the four investigated genes have high expression level in leaf. A gene-to-metabolite network was constructed to reveal the correlation between the gene expression and the metabolite accumulation. All the four investigated genes showed high correlation with PPD type saponins Rb_2, Rb_3 and Rc. Combining metabolic profiling, RT-qPCR and gene-to-metabolite network, the results indicated that the leaf of *P. notoginseng* were the main biosynthesis site of PPD type saponins Rb_2, Rb_3 and Rc. The contribution to the biosynthesis of ginsenoside Rb_2, Rb_3 and Rc was in order of *PnSE* > *PnDS* > *PnSQS* > *PnFPS*. This is the first report integrating the gene expression levels and metabolites accumulation in *P. notoginseng*. The gene-to-metabolite network constructed can be used to screen the key genes of triterpene saponins biosynthesis pathway and provide valuable information to improve the yield of triterpene saponins by plant metabolic engineering in *P. notoginseng*.

6. Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No. 81403043), Natural Science Foundation of Jiangsu Province of China (Grant No. BK20140663), National Natural Science Foundation of Innovative Research Groups (Grant No. 81421005) and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

7. Reference

[1] Wang, C.Z., McEntee, E., Wicks, S., Wu, J.A., and Yuan, C.S. (2006) Phytochemical and analytical studies of *Panax notoginseng* (Burk.) F.H. Chen. J. Nat. Med. **60**, 97-106.

[2] Ng, T.B. (2006) Pharmacological activity of sanchi ginseng (*Panax notoginseng*). J. Pharm. Pharmacol. **58**, 1007-1019.

[3] Xia, P.G., Li, J.Z., Wang, R.L., Zhang, Y., Guo, H.B., Yan, X.J., Liu, Y., and Liang, Z.S. (2015) Comparative study on volatile oils of four Panax genus species in Southeast Asia by gas chromatography-mass spectrometry. Ind. Crop. Prod. **74**, 478-484.

[4] Gao, B., Huang, L., Liu, H., Wu, H., Zhang, E., Yang, L., Wu, X., and Wang, Z. (2014) Platelet P2Y(1)(2) receptors are involved in the haemostatic effect of notoginsenoside Ft1, a saponin isolated from *Panax notoginseng*. Br. J. Pharmacol. **171**, 214-223.

[5] He, N.W., Zhao, Y., Guo, L., Shang, J., and Yang, X.B. (2012) Antioxidant, antiproliferative, and pro-apoptotic activities of a saponin extract derived from the roots of *Panax notoginseng* (Burk.) F.H. Chen. J. Med. Food **15**, 350-359.

[6] Mao, Q., Li, Y., Li, S.L., Yang, J., Zhang, P.H., and Wang, Q. (2014) Chemical profiles and anticancer effects of saponin fractions of different polarity from the leaves of *Panax notoginseng*. Chin. J. Nat. Med. **12**, 30-37.

[7] Yoshikawa, M., Morikawa, T., Kashima, Y., Ninomiya, K., and Matsuda, H. (2003) Structures of new dammarane-type triterpene saponins from the flower buds of *Panax notoginseng* and hepatoprotective effects of principal ginseng saponins. J. Nat. Prod. **66**, 922-927.

[8] Chan, P., Thomas, G.N., and Tromlinson, B. (2002) Protective effects of trilinolein extracted from *Panax notoginseng* against cardiovascular disease. Planta Med. **68**, 1024-1028.

[9] Liu, X.X., Wang, L., Chen, X.Q., Deng, X.T., Cao, Y., and Wang, Q. (2008) Simultaneous quantification of both triterpenoid and steroidal saponins in various Yunnan Baiyao preparations using HPLC-UV and HPLC-MS. J. Sep. Sci. **31**, 3834-3846.

[10] Zhang, H., and Cheng, Y. (2006) Solid-phase extraction and liquid chromatography electrospray mass spectrometric analysis of saponins in a Chinese patent medicine of formulated *Salvia miltiorrhizae* and *Panax notoginseng*. J. Pharm. Biomed. Anal. **40**, 429-432.

[11] Zhang, Z.L., Wang, W.Q., Wang, Y., Yang, J.Z., and Cui, X.M. (2010) Influence of *Panax notoginseng* continuous cropping on seed germination and seedling growth of the plant. Chin. J. Ecology **29(8)**, 1493-1497.

[12] Liang, Y.L., Wu, J.J., Li, Y., Li, J., Ouyang, Y., He, Z., and Zhao, S.J. (2015) Enhancement of ginsenoside biosynthesis and secretion by Tween 80 in *Panax ginseng* hairy roots. Biotechnol. Appl. Biochem. **62**, 193-199.

[13] Chen, L., Lan X.W., Li, S., Zhu H., and Wu, Y.S. (2006) Genetic cloning and sequence analysis of farnesyl pyrophosphate synthase in *Panax notoginseng*. Chin. Tradit. Herbal Drugs **31(7)**, 1080-1083.

[14] Ding, Y.X., Ou-Yang, X., Shang, C.H., Ren, A., Shi, L., Li, Y.X., and Zhao, M.W. (2008) Molecular cloning, characterization and differential expression of a farnesyl-diphosphate synthase gene from the basidiomycetous fungus *Ganoderma lucidum*. Biosci. Biotechnol. Biochem. **72**, 1571-1579.

[15] Niu, Y., Luo, H., Sun, C., Yang, T.J., Dong, L., and Huang, L. (2014) Expression profiling of the triterpene saponin biosynthesis genes *FPS*, *SS*, *SE*, and *DS* in the medicinal plant *Panax notoginseng*. Gene **533(1)**, 295-303.

[16] Wu, Y.S., Zhu, H., Li, S., Zhao, R.Q., and Lan, X.W. (2007) Transcription expression of squalene synthase gene in Root, Stem and Rootstock of *Panax notoginseng* and synthesis of triterpenoids. Chin. J. Biochem. Mol. Biol. **23(12)**, 1000-1005.

[17] He, F., Zhu, Y., He, M., and Zhang, Y. (2008) Molecular cloning and characterization of the gene encoding squalene epoxidase in *Panax notoginseng*. DNA Seq. **19**, 270-273.

[18] Phillips, D.R., Rasbery, J.M., Bartel, B., and Matsuda, S.P. (2006) Biosynthetic diversity in plant triterpene cyclization. Curr. Opin. Plant Biol. **9**, 305-314.

[19] Han, J.Y., Kim, H.J., Kwon, Y.S., and Choi, Y.E. (2011) The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol. **52**, 2062-2073.

[20] Han, J.Y., Hwang, H.S., Choi, S.W., Kim, H.J., and Choi, Y.E. (2012) Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. Plant Cell Physiol. **53(9)**, 1535.

[21] Luo, H.M., Sun, C., Sun, Y.Z., Wu, Q., Li, Y., Song, J.Y., Niu, Y.Y., Cheng X.L., Xu, H.X., Li, C.Y., Liu, J.Y., Steinmetz, A., and Chen, S.L. (2011) Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. BMC Genomics **12(Suppl. 5)**, S5-S19.

[22] Yue, C.J., Zhou, X. and Zhong, J.J. (2008) Protopanaxadiol 6-Hydroxylase and its role in regulating the ginsenoside heterogeneity in *Panax notoginseng* cells. Biotechnol. Bioeng. **100(5)**, 933-940.

[23] Yue, C.J. and Zhong, J.J. (2005a) Impact of external calcium and calcium sensors on ginsenoside Rb1 biosynthesis by *Panax notoginseng* cells. Biotechnol. Bioeng. **89**, 444-452.

[24] Yue, C.J. and Zhong, J.J. (2005b) Purification and characterization of UDPG: ginsenoside Rd glucosyltransferase from suspended cells of *Panax notoginseng*. Process Biochem. **40**, 3742-3748.

[25] Kim, Y.J., Zhang, D.B., and Yang, D.C. (2015) Biosynthesis and biotechnological production of ginsenosides. Biotechnol. Adv. **33**, 717-35.

[26] Lai C.J.S., Tan T., Zeng S.L., Dong X., Liu E.H., and Li P. (2015) Relative quantification ofmulti-components in *Panax notoginseng* (Sanqi) by high-performance liquid chromatography with mass spectrometry using mobile phase compensation. J. Pharm. Biomed. Anal. **102**,150-156.

[27] Li, F., Cheng, T.F., Dong, X., Li, P., and Yang, H. (2016) Global analysis of chemical constituents in Shengmai injection using high performance liquid chromatography coupled with tandem mass spectrometry. J. Pharmaceut. Biomed. **117**, 61-72.

[28] Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $-^{\Delta CT}$ method. Methods **25**, 402-408.

[29] Chen, R., Li, Q., Tan, H., Chen, J., Xiao, Y., Ma, R., Gao, S., Zerbe, P., Chen, W., and Zhang, L. (2015) Gene-to-metabolite network for biosynthesis of lignans in MeJA-elicited *Isatis indigotica* hairy root cultures. Front. Plant Sci. **6**, 952.

[30] Xiao, Y., Gao, S., Di, P., Chen, J., Chen, W., and Zhang, L. (2009) Methyl jasmonate dramatically enhances the accumulation of phenolic acids in *Salvia miltiorrhiza* hairy root cultures. Physiol. Plant. **137**, 1-9.

[31] Toh, D.F., New, L.S., Koh, H.L., and Chan, E.C. (2010) Ultra-highperformance liquid chromatography/time-of-flight mass spectrometry (UHPLC/TOFMS) for time-dependent profiling of raw and steamed *Panax notoginseng*. J. Pharm. Biomed. Anal. **52**, 43-50.

[32] Wan, J.B., Yang, F.Q., Li, S.P., Wang, Y.T., and Cui, X.M. (2006) Chemical characteristics for different parts of *Panax notoginseng* using pressurized liquid extraction and HPLC-ELSD. J. Pharm. Biomed. Anal. **41**, 1596-1601.

[33] Ruan, J.Q., Leong, W.I., Yan, R., and Wang, Y.T. (2010) Characterization of metabolism and in vitro permeability study of notoginsenoside R_1 from *Radix notoginseng*. J. Agric. Food Chem. **58(9)**, 5770-5776.

[34] Kim, D.H. (2012) Chemical Diversity of *Panax ginseng*, *Panax quinquifolium*, and *Panax notoginseng*. J. Ginseng. Res. **36(1)**, 1-15.

[35] Rischer. H., Oresic, M., Seppanen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C., Inze, D., Oksman-Caldentey, K.M., and Goossens, A. (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc. Natl. Acad. Sci. USA **103**, 5614-5619.

Table 1 Content of twelve triterpene saponins in different parts of one-year-old *P. notoginseng*

No.	analytes	Leaf	Petiole	Root
PPD				
5	ginsenoside Rb_1	3.30±0.24	3.49± 0.59	2.33±0.73
9	ginsenoside Rb_2	8.92± 0.57	0.55±0.14	0.15± 0.04
8	ginsenoside Rc	9.80±0.75	0.19±0.06	0.31±0.13
11	ginsenoside Rd	0.57±0.11	0.34±0.09	0.54±0.27
10	ginsenoside Rb_3	21.52±1.07	0.47±0.15	1.14±0.27
12	ginsenoside Rg_3	nd	nd	nd
PPT				
1	notoginsenoside R1	nd	nd	1.28±0.52
2	ginsenoside Rg_1	nd	nd	5.34±1.25
3	ginsenoside Re	0.09±0.02	0.17±0.05	2.73±0.32
4	ginsenoside Rf	nd	nd	nd
6	ginsenoside Rg_2	nd	nd	0.50±0.05
7	ginsenoside Rh_1	nd	nd	0.15±0.05
	PPD	44.11±2.74	5.04±1.03	4.47±1.44
	PPT	0.09±0.02	0.17±0.05	10.00±2.19
	Total	44.20±2.76	5.21±1.08	14.47±3.63
	PPD/PPT	490	29.65	0.447

Note: The unit of data is mg/g, average data with standard errors from three independent biological experiments are shown above, nd means undetected.

Table 2 Primers used for RT-qPCR in this study

Primers	*Primer Sequence(5'-3')*
PnFPS-RT-F	**CTCAAGAAGCATTTCCGACAAAAGCCTTAC**
PnFPS-RT-R	**ACAAGTCTTTCTCTCCTACAAGGGTGG**
PnSQS-RT-F	**GAGCACTTGACACTGTTGAGGATGAC**
PnSQS-RT-R	**CTATTGCCTCCTGGTAACCGTTTCC**
PnSE-RT-F	**GGTGAACTTCTACAACCAGGAGGCTAC**
PnSE-RT-R	**CTCCAAGGGGTAAGAAAGCCTGGTGTT**
PnDS-RT-F	**CAAATAATGAGTTGGTGGGCAGAAGAT**
PnDS-RT-R	**TGGTGGCGATAATTGCTTGAGTAGC**
PnActin-RT-F	**TCGGACAACGAGGCAGCACTTT**
PnActin-RT-R	**GCTAAAGAGCAGCCAACAGGCC**

Figure captures

Fig. 1 Biosynthetic pathway of triterpene saponins in *P. notoginseng*. HMG-CoA, hydroxymethylglutaryl CoA; HMGR, HMGR-CoA reductase; MVA, mevalonic acid; G3P, glyceraldehyde-3-phosphate; DXS, DXP synthase; DXP, deoxyxylulose 5-phosphate; DXR, DXP reductoisomerase; MEP, methylerythritol 4-phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; FPS, farnesyl diphosphate synthase; SQS, squalene synthase; SE, squalene epoxidase; DS, dammarenediol II synthase; PPDS, protopanaxadiol synthase; P6H(PPTS), protopanaxadiol 6-hydroxylase (protopanaxatriol synthase); UDP-glycosyltransferases, UGTs; glycosyltransferases, GTs; UGRdGT, UDP-glucose:ginsenoside Rd glucosyltransferase.

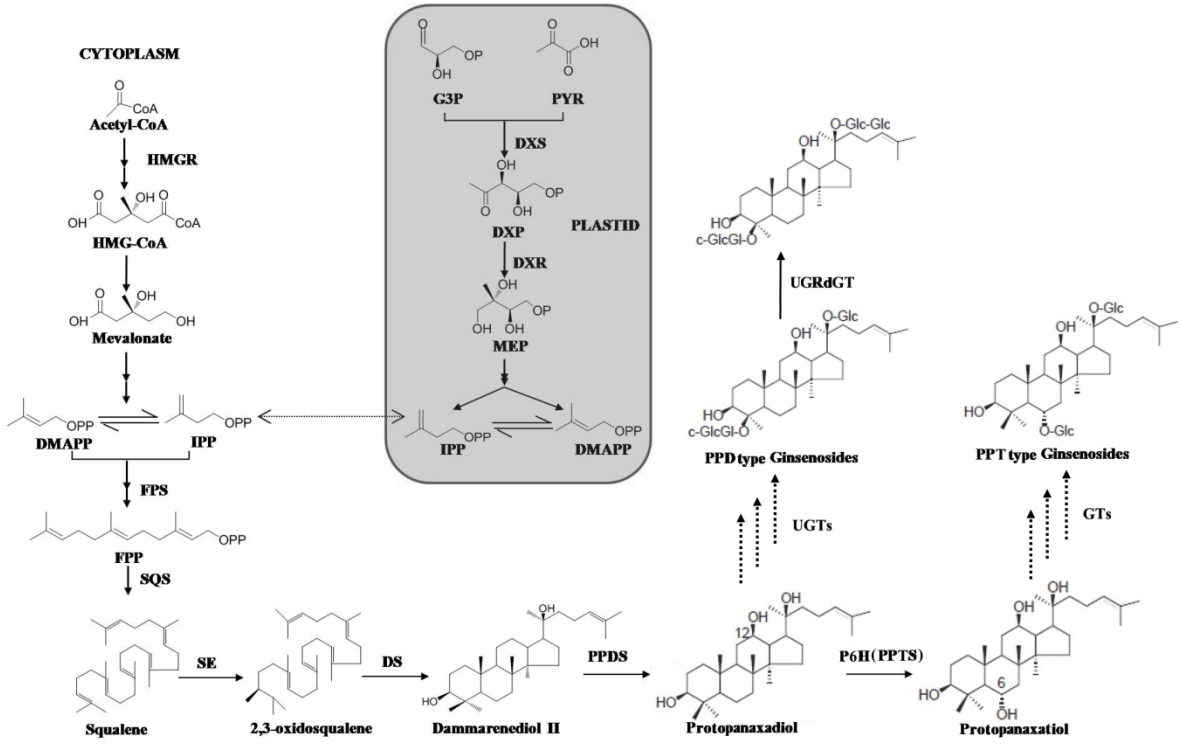

Fig. 2 The selected ion monitoring chromatograms of (A) leaf (B) petiole and (C) root from one-year-old *P. notoginseng* by HPLC-MS in negative ion mode. 1: notoginsenoside R1; 2: ginsenoside Rg_1; 3: ginsenoside Re; 4: ginsenoside Rf; 5: ginsenoside Rb_1; 6: ginsenoside Rg_2; 7: ginsenoside Rh_1; 8: ginsenoside Rc; 9: ginsenoside Rb_2; 10: ginsenoside Rb_3; 11: ginsenoside Rd; 12: ginsenoside Rg_3; IS: jujuboside B.

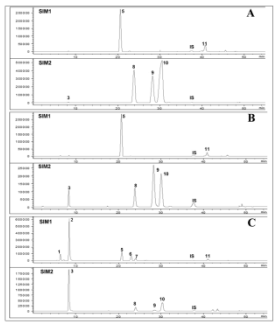

Fig. 3 (A) PCA scores plot of the leaf, petiole and root in one-year-old *P. notoginseng*. Square represents root, triangle represents petiole, and rhombus represents leaf. (B) The loading plot of PCA (p1/p2).

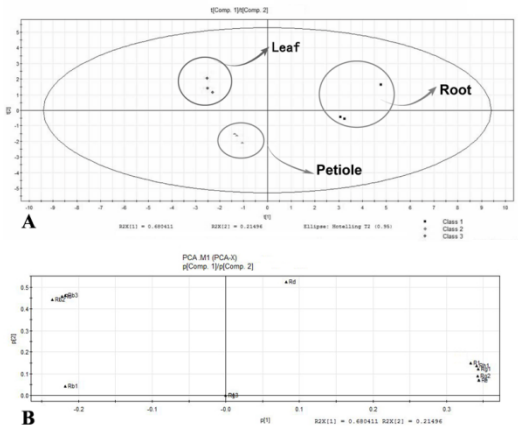

Fig. 4 Expression patterns of *PnFPS*, *PnSQS*, *PnDS* and *PnSE* in the root, petiole and leaf of one-year-old *P. notoginseng*. Average data with standard errors from three independent biological experiments are shown. *PnActin* is used as a control for normalization.

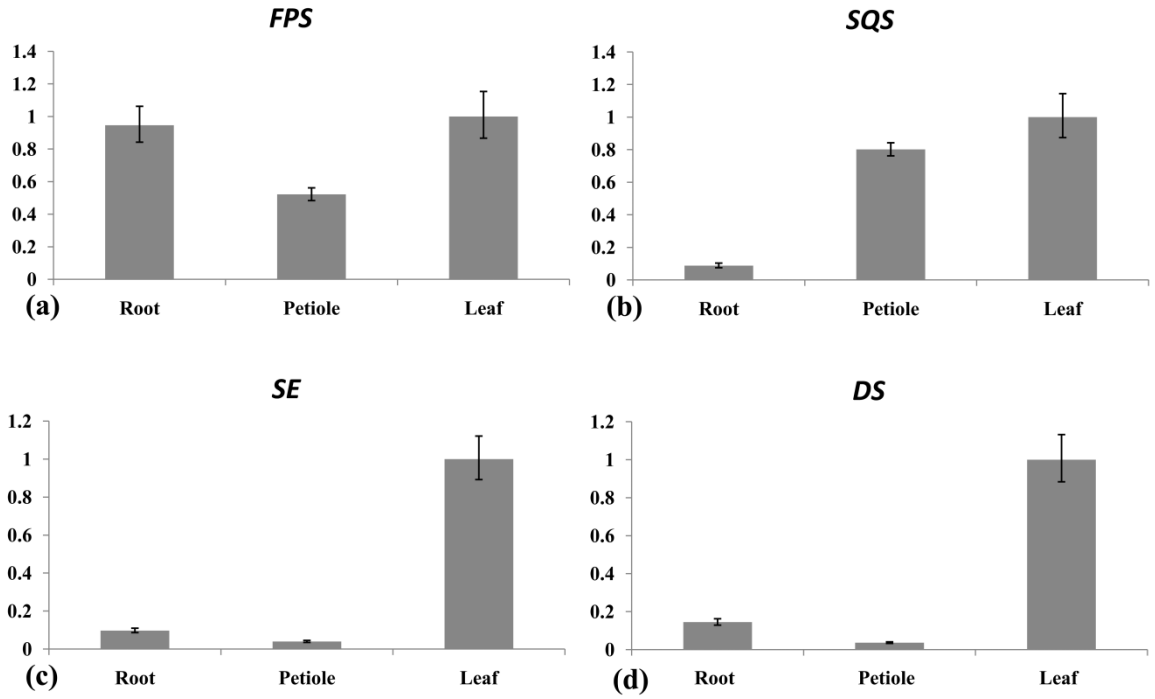

Fig. 5 Gene-to-metabolite network for one-year-old *P. notoginseng*. Squares on the left and right represent genes and metabolites, respectively. The canonical correlation coefficient between two canonical correlation variables (U and V) was 1. Dotted lines denoted the correlation coefficient between raw variables (gene or metabolite) and canonical correlation variables (U or V). Variable correlation coefficient between genes and metabolites are illustrated as the numbers on the corresponding lines.